



King's Research Portal

DOI:

[10.1177/2053951714543804](https://doi.org/10.1177/2053951714543804)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Venturini, T., Baya Laffite, N., Cointet, J.-P., Gray, I., Zabban, V., & De Pryck, K. (2014). Three maps and three misunderstandings: A digital mapping of climate diplomacy. *Big Data & Society*, 1(2).
<https://doi.org/10.1177/2053951714543804>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Three maps and three misunderstandings: A digital mapping of climate diplomacy

Tommaso Venturini¹, Nicolas Baya Laffite¹,
Jean-Philippe Cointet², Ian Gray¹, Vinciane Zabban³
and Kari De Pryck¹

Abstract

This article proposes an original analysis of the international debate on climate change through the use of digital methods. Its originality is twofold. First, it examines a corpus of reports covering 18 years of international climate negotiations, a dataset never explored before through digital techniques. This corpus is particularly interesting because it provides the most consistent and detailed reporting of the negotiations of the United Nations Framework Convention on Climate Change. Second, in this paper we test an original approach to text analysis that combines automatic extractions and manual selection of the key issue-terms. Through this mixed approach, we tried to obtain relevant findings without imposing them on our corpus. The originality of our corpus and of our approach encouraged us to question some of the habits of digital research and confront three common misunderstandings about digital methods that we discuss in the first part of the article (section ‘Three misunderstandings on digital methods in social sciences’). In addition to reflecting on methodology, however, we also wanted to offer some substantial contribution to the understanding of UN-framed climate diplomacy. In the second part of the article (section ‘Three maps on climate negotiations’) we will therefore introduce some of the preliminary results of our analysis. By discussing three visualizations, we will analyze the thematic articulation of the climatic negotiations, the rise and fall of these themes over time and the visibility of different countries in the debate.

Keywords

Digital methods, climate negotiations, climate change, science and technology studies, text analysis, network analysis

Introduction

This article proposes an original analysis of the international negotiations on climate change through the use of digital methods. Its originality is twofold.

First, it examines a corpus of reports on climate negotiations never explored before through digital techniques. This corpus is particularly interesting because it provides the most consistent and detailed reporting of the proceedings of the negotiations of the United Nations Framework Convention on Climate Change (UNFCCC) between 1995 and 2013. Though digital data have been already employed to analyze the discussions on global warming, their use has been so far limited to more “traditional” datasets extracted from bibliographic archives (see, for example, Anderregg et al., 2010; Bjurström and Polk, 2011; Li et al., 2011;

Stanhill, 2001) and social media (Niederer, 2013; Rogers and Marres, 2000). Compared to these “traditional” datasets, our corpus is novel for its conditions of production – the reports are digital summaries of roundtable discussions of the UNFCCC meetings – and for its challenging format – each summary compiles statements made by all the participating parties.

Second, in this paper we test an original approach to text analysis that combines automatic extractions and

¹Sciences Po, médialab, Paris, France

²INRA, Marne-la-Vallée, France

³LATTS, Marne-la-Vallée, France

Corresponding author:

Tommaso Venturini, Sciences Po, 27 rue St Guillaume, 75337 Cedex 07, Paris 75337, France.

Email: tommaso.venturini@sciences-po.fr



manual selection of the key issue-terms. Through such an approach, we wanted to avoid the drawbacks of grounding the extraction of the terms uniquely on quantitative indicators or on the qualitative judgment of experts. Iterating the interactions between expertise and computation, we let the dictionary of analysis emerge from our texts, but not in a purely automatic way. Through this mixed approach, we tried to obtain relevant findings, without imposing them on our corpus.

The originality of our corpus and of our approach encouraged us to question some of the habits of digital research and confront three common misunderstandings about digital methods that we discuss in the first part of the article (section ‘Three misunderstandings on digital methods in social sciences’).

In addition to reflecting on methodology, however, we also wanted to offer some substantial contribution to the understanding of UN-framed climate diplomacy. In the second part of the article (section ‘Three maps on climate negotiations’) we will therefore introduce some of the preliminary results of our analysis. By discussing three visualizations, we will analyze the thematic articulation of the climatic negotiations, the rise and fall of these themes over time and the visibility of different countries in the debate.

Three misunderstandings on digital methods in social sciences

The appearance of digital traces has triggered contradictory feelings among social scientists. While optimists welcome enthusiastically a new source of information, potentially richer and less expensive than traditional datasets (Kleinberg, 2008; Watts, 2007), pessimists show distrust towards the turmoil caused by a flood of muddled traces.

By this article, we wish to contribute to the debate on digital data in the social sciences (cf. Giles, 2012; Lazer et al., 2009) by proposing an approach that is neither enthusiastic nor alarmist. Most difficulties, we believe, derive from a misunderstanding on the nature of digital methods, or, to be more precise, from three misunderstandings: first, a split-conception of digital traces that is both too restrictive and too ambitious; second, a vacillation between disregard and distrust on the conditions of production of digital traces; finally, a tendency to mistake “digital” for “automatic.” Though often uttered orally, these misunderstandings seldom appear in published papers (scholars who dislike digital methods tend to avoid them more than explicitly criticizing them). Direct or reported evidence of such misunderstandings can, however, be found in literature (Barnes, 2014; Bollier and

Firestone, 2010; Boyd and Crawford, 2011; Carr, 2014; Cresswell, 2014; Dalton and Thatcher, 2014; Gayo-Avello, 2012; Harford, 2014; Law et al., 2011; Lazer et al., 2014; Manovich, 2012; Marcus and Davis, 2014; Marres, 2012; Rieder and Röhle, 2012; Savage and Burrows, 2007; Uprichard, 2012, 2013). In the first part of this article, we will address these three misunderstandings by describing how we came across them in our work on climate negotiations.

First misunderstanding: Which digital traces?

The creation of a “diplomatic” corpus. Adopted in 1992 at the Earth Summit,¹ enforced in March 1994, and ratified by 195 countries, the UNFCCC aims at

stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system [...] within a time frame sufficient to allow ecosystems to adapt naturally to climate change, to ensure that food production is not threatened and to enable economic development to proceed in a sustainable manner.²

Although UNFCCC’s discussions are often documented by NGOs, political constituencies, and researchers, these traces are not systematically collected and easily exploitable by digital research. Therefore, we first had to identify traces sufficiently representative and sufficiently structured to be processed by the available tools of analysis.

Among the available documentation, the most interesting is Volume 12 of the *Earth Negotiations Bulletin* (ENB).³ Released by the International Institute for Sustainable Development (IISD), the ENB provides daily reports on 33 separate UN negotiations on environment and development. The ENB archive contains thousands of reports on dozens of multilateral agreements. Volume 12, dedicated to climate negotiations, contains 594 issues. The ENB provides a workable proxy with which to analyze all the climate negotiations (and avoid drowning in the thousands of disparate documents stocked on the website of the UNFCCC). In this article, we will principally recount the construction and treatment of this dataset that is so much unlike the sources generally used by practitioners of digital methods.

There are more things in Internet and Web (Horatio) than are indexed in your platforms. The necessity to use very specialized traces such as those provided by the ENB led us to reflect on the first misunderstanding on digital methods: the supposed impossibility to standardize results obtained on digital samples to offline populations

(Couper, 2000).⁴ This difficulty is, to a large extent, due to the fact that digital traceability is often reduced to its most visible instances: the huge Internet platforms such as Google, Facebook, Wikipedia, Twitter, LinkedIn, or Amazon. Too often, the myth of “big data” reduces the richness of digital data to a mere question of size. Hence, the delusion that the mere size of social media’s APIs (Application Programming interface) could be the key to understanding collective life. Unfortunately, as classic statistics has made very clear, the representativeness of a sample depends only indirectly upon its size. The quality of a sample rather lies in its similarity to the sampled population and in its capacity to include the same variability:

Typically researchers focus on sample size as the most important consideration in achieving representativeness: how many texts must be included in the corpus, and how many words per text sample. Books on sampling theory, however, emphasize that sample size is not the most important consideration in selecting a representative sample; rather, a thorough definition of the target population and decisions concerning the method of sampling are prior considerations. Representativeness refers to the extent to which a sample includes the full range of variability in a population. (Biber, 1993: 243)

The value of the digital traceability is not a question of size and cannot be measured by the gigantism of a few digital platforms. The real impact of the digital traceability should be sought in the increasing *diversity* of the traces available for social research. This diversification allows quantifying objects that so far were hard to measure and encourages us to reflect on how we select and build our corpora. Such is the first methodological lesson we draw from the difficulties and the possibilities offered by our “unconventional” diplomatic corpus.

Second misunderstanding: Whose digital traces?

It is important to clarify the meanings of two expressions frequently used in this article: “digital traces” and “digital data.” Though there are many ways of defining both terms (see Reigeluth, 2014 for a detailed discussion), we propose here a simple and methodological distinction. In this article, “digital traces” refers to any set of bits stored in the memory of a digital device (a computer, in most cases) as a result of the deliberate implementation of tracing systems. “Digital data” are instead the organized set of information, produced from digital traces through the work of researchers that select, clean and exploit them in a specific study. Our distinction between traces and data is deliberately *non-substantive*: traces are not necessarily more

raw, spurious, or primitive than data. The difference is purely *methodological*. Its objective is to remind us that, in digital research, the work done to adapt the information to the investigation’s objectives (data) is most of the time preceded by the work done by someone else to make such information available (traces).⁵ The trace/data distinction invites us to consider both operations with the greatest attention as they both influence research results.

Our work on the ENB corpus provides a good example of this, as knowledge of its production context was essential not only in the harvesting process but also throughout the whole analysis and the interpretation of the results. The second lesson we can draw from our case study concerns therefore the origins of digital traces and encourages us to reflect on the processes that convert traces into data.

The building up of the ENB corpus and its original recipients. The ENB was started as an initiative of three experts participating in the Rio Summit of 1992. At the end of the conference, the IISD contacted the founding members and offered to publish the report during the next negotiations. Today, the writing of the ENB report results from a collective work of four permanent experts,⁶ several part-time administrative staff, two full-time translators, and 60 consultant experts from 32 countries. Most of the members are PhD candidates or PhDs with some experience in the domain of environment and development.

In the case of the climate negotiations, ENB covers the official proceedings of conferences and, when possible, the discussion in the corridors. Each ENB issue reports on a day of negotiations. It includes an introduction, a short history of the negotiations, and the transcription of the discussions. The first audience for these texts is the negotiations’ actors, to whom paper reports are distributed during the negotiations. The complete archive of the issues is then available on the IISD website. The ENB reports do not go into the details of the discussions, but they propose a point-by-point paraphrasing of the arguments exchanged and a summary of the outcomes of the negotiations.

The content of the reports must be balanced, independent, and substantiated by the testimony of several participants. In order to ensure consistency in style, ENB authors and editors are specifically trained and provided with a style handbook. This handbook details the structure of the reports (characterized by very strict *templates*: establishing, for example, a maximum length of 1800–1900 words for the *daily issues*). The handbook prescribes a clinical style – warning the authors against adjectives and adverbs: “while it makes writing and reading the Bulletin particularly boring, it also removes

any overt signs of subjectivity” (*ENB Manual and Style Guide*, p. 53) and suggests a list of 65 preferable verbs to avoid injecting ambiguity or motive regarding actors’ statements.

The very standardized structure of the ENB is one of its main interests as a collection for digital analysis: unlike the UNFCCC original negotiation documents, this high level of standardization (in terms of language, text structure, availability of metadata) makes this corpus relatively homogeneous and thus facilitates automatic processing.

(Most) digital data were not created for social research. Exploiting a set of traces collected for purposes other than scientific research is not unusual digital research. Most digital traces are collected for purposes of marketing (such as loyalty or credit cards), surveillance (as in air travel), technical optimization (as in telecommunication networks), or information sharing (as in the ENB reports we address in this paper). In one way or the other, they are *second-hand* data, the production of which is not directly controlled by the researchers.

Using these traces requires therefore questioning the conditions of their production. When the World Bank published its data (openknowledge.worldbank.org/about), we have to ask how these data were computed and why they are disclosed (Tabor, 2012). If American Online releases by mistake 20 million search requests,⁷ we have to ask whether the use of these data is ethically fair (Ess and AoIR ethics working committee, 2002). If Wikipedia grants access to the full edit history of all its articles (mediawiki.org/wiki/API), we have to reflect on the epistemic statute of this collective enterprise (Viegas et al., 2007).

To reflect on how digital data were created means to resist the temptation to naturalize them. Digital data are more similar to traces collected in a bubble chamber than to footprints left on wet sand: they exist because someone collected and processed them. Remaining conscious that digital data are always produced in a specific context, however, does not lessen their interest. Despite the etymology, data are never given: data always result from a long chain of actions, some of which escape the control of the experimenter (Latour, 1995). To admit that digital traces are not natural items but artifacts created in a specific environment and with specific objectives does not reduce their value. The process of their creation also bears relevant knowledge.

In our case, the ENB writing process is particularly well documented, and this oriented our analysis: the standardization of the texts makes lexical analysis relevant; the thematic focus of paragraphs defines the granularity of the co-occurrences of words; the use of

a limited list of verbs allows us to identify the position of the actors, etc.

Third misunderstanding: What is digital research?

The abundance of digital data and computational resources has stirred a scientific excitement over a potential “mechanization” of digital research. “Big data” would reduce the need for the researcher’s intervention, evacuate the task of fitting data to existing theories, and allow prediction without the fragility of interpretation. In other words, the researchers’ intervention would become increasingly expendable. The confidence in quantification is not recent (think of Isaac Asimov’s “psychohistory”), but the development of digital technology has renewed this faith.

Our experience with the ENB corpus, however, points to different conclusions. Even though without the help of the computer it would have been impossible to consider each and every word of the ENB Volume 12 (given its size and complexity), the intervention of the researcher remains crucial. Indeed, simply running the raw material through the “machine” could only produce disappointing results. But the point here is not to defend manual analyses against automated ones (or the other way around), rather to show that good results can be obtained only by combining the two.

The slow process of map creation. Our corpus was built from the 594 issues contained in Volume 12 of the ENB, from the conferences in New York in 1995 to the conference in Warsaw in 2013. From all these issues, we only kept the *daily issues* related to the COPs (Conference of the Parties).⁸ To keep close to the diplomatic dimension of the negotiations, we put aside the issues related to other UNFCCCS bodies such as the Subsidiary Body for Scientific and Technological Advice and the Subsidiary Body for Implementation.

The second step is dedicated to analysis. Ours relies on paragraphs – each paragraph corresponding to a natural thematic unit according to the production and format of these reports. The corpus analysis platform Cortext (<http://manager.cortext.net/>) of IFRIS was then used to (1) analyze the lexical content of the collection of bulletins, (2) identify the main clusters emerging from the negotiations thanks to the network analysis applied to a matrix of co-occurrences between extracted words, and (3) analyze the profiles of different actors by qualifying which frame of words was mobilized by which countries during different COPs.

The analysis starts by the lexical extraction (using a mixed algorithm constructed around linguistic and statistical approaches) of the most relevant nominal groups (which will simply be referred to as “terms” for more convenience) in the whole set of corpus

paragraphs (5663). Since the algorithms are used to sort the terms according to their frequency and “specificity,” by force, they pick up a range of uninteresting or irrelevant terms (as well as mis-categorizing relevant terms); we cleaned and corrected the term lists manually, which allowed us to keep only the terms referring to well-identified themes (“issues”) that could be qualified as actual topics under discussion in the negotiations. This process involved a great deal of iteration, requiring us to go back and forth between the texts of the corpus itself, the data, the statistics connected to the lexical extraction, and the network of co-occurrences.

Finally, we have kept terms characterized by their frequency (at least seven occurrences in the corpus) and their relevance (the tendency to be used in specific linguistic contexts of the negotiations, e.g. “historical responsibility”). We have rejected terms that were not specific enough or could be ambiguous (e.g. “increased efforts”). We have also merged the declensions of terms (“social cost” and “social costs”), equivalent forms (e.g. “technological transfer” and “transfer of technologies”), and terms that were clearly synonyms (e.g. GHGs and Green House Gases). Out of a list of 1178 terms extracted, we have thus compiled a dictionary of terms of about 300 nominal groups (both lists are available at: <http://medialab.sciences-po.fr/publications/misunderstandings/>).

Finally, we have “mapped” the negotiations’ topics and created, from this map, thematic profiles. The operations we carried out to render the results of our analysis visually, through networks and stream-graphs, are detailed below and are crucial as they distinguish our investigation from more conventional approaches of computer-assisted discourse analysis.

The count of term co-occurrences within one paragraph of the corpus allowed us to calculate a semantic network. In this network, terms are linked with a strength proportional to a co-occurrence-based measure of similarity introduced by Weeds and Weir (2005). In brief, two terms are all the more close when they co-occur with the same terms. This distributional measure builds a semantic weighted network featuring a very large number of edges that should be filtered before being mapped. We rely on the topological properties of the network to eliminate those links whose intensity is below a threshold. We simply find the critical parameter for which the network is still made of one giant connected component. The rationale behind this choice is that we want to preserve the “macroscopic structure” of the network (how different parts of the network are relatively positioned) while selecting only most salient edges that contribute to the mesoscopic structures. A *clustering* algorithm (Blondel et al., 2008) then allowed us to identify 12 cohesive subsets

of the network corresponding to the main themes discussed in the negotiations (see Figure 1). Each paragraph in the corpus has then been labeled with one or several topics from this list, according to the terms it contains (see Figure 10). Paragraphs have also been labeled according to the “players” referred to in the text (at least one occurrence of the name or demonym of the country involved) (see Figure 7).

Digital is not automated. The last misunderstanding on digital methods is both the subtlest and the most difficult to overcome. The difficulty lies in a double ambiguity that first takes “digital” for “automated” and then “automated” for “objective.” To correct this misunderstanding, we need to tackle these two parts separately.

The first ambiguity comes from a tendency to see in digital technologies nothing but the capacity to automate repetitive processes. The computer then would only be a tool to relieve researchers from their most “mechanical” tasks. This is undoubtedly one of the computer’s contributions, but certainly not the only one. The history of sciences teaches that whenever a new tool imposes itself in research practice, it influences the course of that science. The telescope has not only allowed us to see farther but it has also allowed us to see differently. The printing press did not only allow us to print more books, but it also allowed us to print new texts (Eisenstein, 1979). Thus, digital technologies do not just assist traditional research methods; they create new scientific practices (Rogers, 2013).

Researchers willing to make their life easier would not find here what they are looking for. Digital methods question research habits and assumptions and require taking many decisions. We are very far from the concept of automation: computerized research is neither faster nor easier. The experience of a digital project is the experience of a series of successive impediments. First, choices must be made on how to harvest the traces: what we are tracking is not as easy to collect as we had thought. The traces are messier than expected, and transforming them into data is problematic (how to correct the mistakes, detect the duplicates, manage normalization, remove irrelevant results?). Second, we have to select the most appropriate analysis tools: analysis algorithms are numerous, but they are all poorly documented and must all be adjusted to the available data. Third, results must be visualized, but how to choose among dozens of possible results and visualizations? And after having overcome all these impediments, we often obtain disappointing results that force us to reconsider all choices made throughout the process. And start again.

But the difficulty does not only lie in the amount of work necessary to practice digital methods.

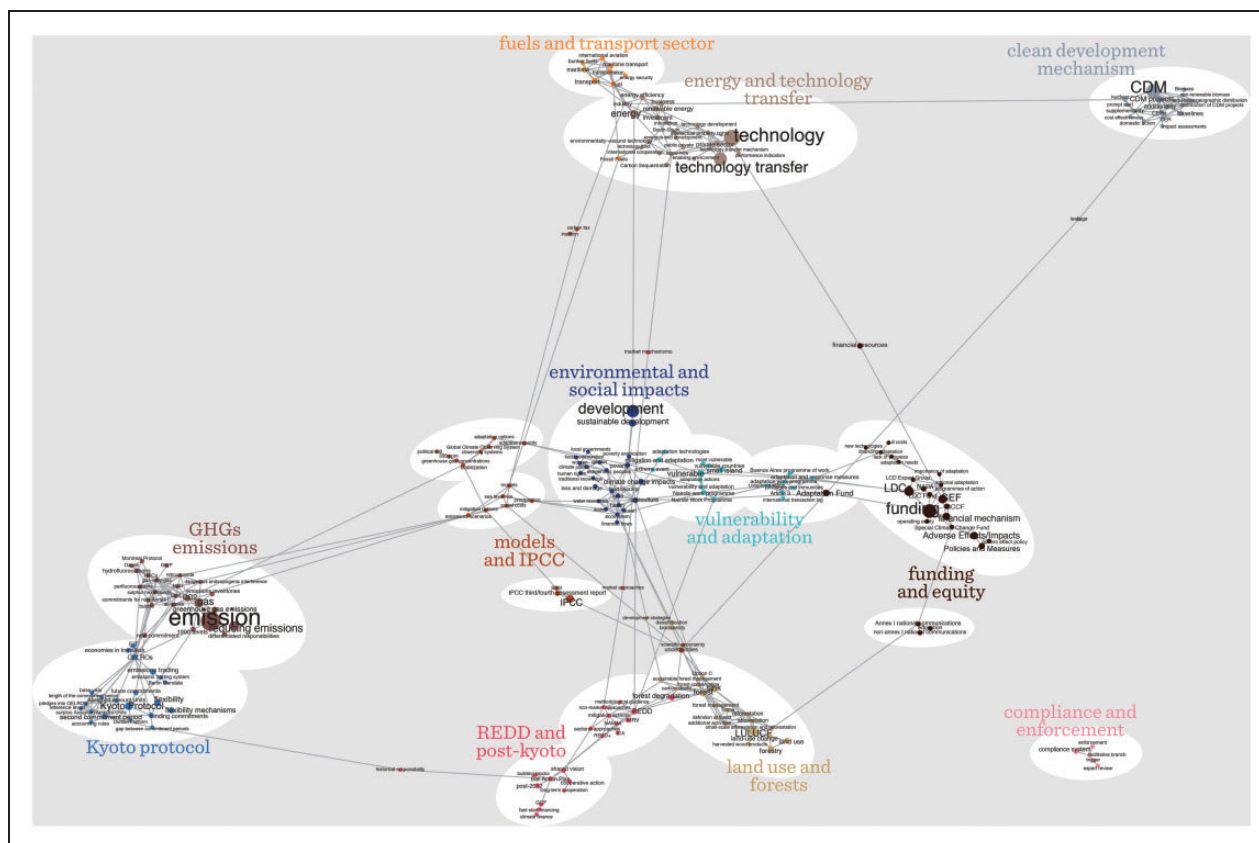


Figure 1. Network of terms co-occurring in the same paragraphs of the ENB. Node position is determined by a force vector algorithm (Jacomy et al., forthcoming) bringing together terms directly or indirectly linked, and keeping away terms with fewer co-occurrences. Node size is proportional to their frequency in the corpus. Node color follows the clusters identified by the clustering algorithm. The names of the clusters have been attributed manually. A high-resolution and zoomable version of this image can be found at: <http://medialab.sciences-po.fr/publications/misunderstandings/> and <http://bds.sagepub.com/content/1/2/2053951714543804/F1.large.jpg>.

The obstacle is often conceptual and results from the second ambiguity hidden in this last misunderstanding: the one between “automated” and “objective.” The Cartesian precept on the evidence of scientific truth lends potent strength to the idea that automation might lead us to an epistemological Promised Land. In this version of digital utopianism, we do not automate research processes out of laziness, but because we want to make results more “objective.” The idea is that the more mechanical the process, the fewer interventions are required from the researcher, the more the results will be free from the researcher’s interpretive subjectivity.

To be sure, the new digital traces encourage us to question the presuppositions of the classic social theories (Latour et al., 2012; Venturini and Latour, 2010) and to observe social phenomena more directly. However, this does not mean that results could emerge spontaneously from the data with no need of human arbitration. This utopia is best described in a provocative article by Chris Anderson apocalyptically titled

“The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”:

Scientists are trained to recognize that correlation is not causation, that no conclusions should be drawn simply on the basis of correlation between X and Y (it could just be a coincidence). Instead, you must understand the underlying mechanisms that connect the two. Once you have a model, you can connect the data sets with confidence. Data without a model is just noise. But faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete...

There is now a better way. Petabytes allow us to say: “Correlation is enough.” We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

It does not take a relentless philosopher, nor a staunch supporter of human thought to find the weakness in this argument. Anybody who has experienced digital methods knows that data are never given. Anderson refers to “numbers,” “measure,” and “correlation” as if they were obvious and unproblematic, whereas experience of computerized research teaches how difficult it is to choose among the proliferation of numbers, measures, and correlations always contradictory and questionable. And the increasing number and diversity of sources brought about by “big data” is not making things any easier. Far from “throwing numbers into big computing clusters,” digital research is about making uncertain and painful arbitrations on the basis of our “non-automatizable” understanding. This suggests, by the way, that (until our methods are more stabilized) it might be a good idea to refrain from using the biggest datasets and focus instead on *small or medium* digital data – which allows direct observation and does not require us to rely exclusively on algorithmic treatment.

Three maps on climate negotiations

It is now time to provide some evidence that the approach we just described can be used to obtain relevant results in researching social phenomena. Drawing on the data of the ENB corpus and navigating our way through the three misunderstandings we discussed, we have produced three “maps” on the international negotiations on climate change: one network of the terms co-occurring in the paragraphs of the ENB and two stream-graphs presenting the variations in the visibility of the different countries and themes of the negotiations. In this second part of the article, we will propose a possible reading of these maps. Such a reading is meant not only to offer insights to climate experts and negotiation actors⁹ but also to provide an example of visual data exploration, which could hopefully inspire scholars working on other digital data.

First map: The negotiations issues

The first map shows a graph of terms related to *issues* discussed throughout climate negotiations since the UNFCCC was adopted (1995–2013) (see Figure 1). There are many ways of making sense of a co-occurrence graph. The most common is to calculate a series of statistics in order to detect the most important terms and their semantic context. In this article, we will propose a different approach, one that is based on the visual observation of the co-occurrence network. We will, in particular, consider one cluster after the other and follow the connections of the most important terms.

The network is organized around five main macro-themes (Figure 1):

- GHGs emissions and Kyoto Protocol (on the left).
- Fuels and transport sector, energy and technology transfers, and clean development mechanism (CDM) (at the top).
- Carbon sinks – Reducing emissions from deforestation and forest degradation (REDD) and land use, land use change and forests (LULUCF) (at the bottom).
- Impacts, vulnerability and adaptation, and funding and equity (on the right).
- Models and Intergovernmental Panel on Climate Change (IPCC) (at the center).

The crossroad of scientific expertise. The cluster “Models and IPCC” (see Figure 2) is located at the intersection of four other themes of the negotiations. To describe the hub role of this cluster, let us take a closer look at the terms of *stabilization* designating the UNFCCC fundamental target. The terms that surround it more closely may serve to characterize it: *greenhouse gas concentrations* (the object of stabilization), *350 parts per million* (the target figure), *carbon dioxide-equivalent* (the measure), and *the political will* (necessary to reach stabilization).

Following the connections of the *stabilization* we can describe how the negotiations are articulated. Toward the left, *stabilization* connects to CO_2 , at the heart of the “GHG emissions measure” cluster (in green) which deals with the issue of greenhouse gases estimation. Toward the bottom, *stabilization* is connected to the *mitigation options*, which is one of the two main strategies to fight against climate change. Toward the top, *stabilization* is connected to terms related to the other main strategy to fight climate change: planned and voluntary adaptation to the consequences of climate change.

The mitigation framework. Terms related to the efforts to mitigate climate change (by reducing GHG emissions) characterize 7 of the 12 clusters of the networks. These clusters are grouped in three main semantic arenas widely scattered across the graph.

Reducing emissions: The first mitigation arena is composed of two clusters representing two fundamental aspects of the UNFCCC process: *GHGs emissions* (GHGs) and the *Kyoto Protocol* (see Figure 3). The proximity of these two clusters reveals the primary target of the UNFCCC: stabilize GHGs concentrations by mechanisms to reduce emissions.

In the “GHGs emissions” (green) cluster, we can discern three types of terms. First, there are terms

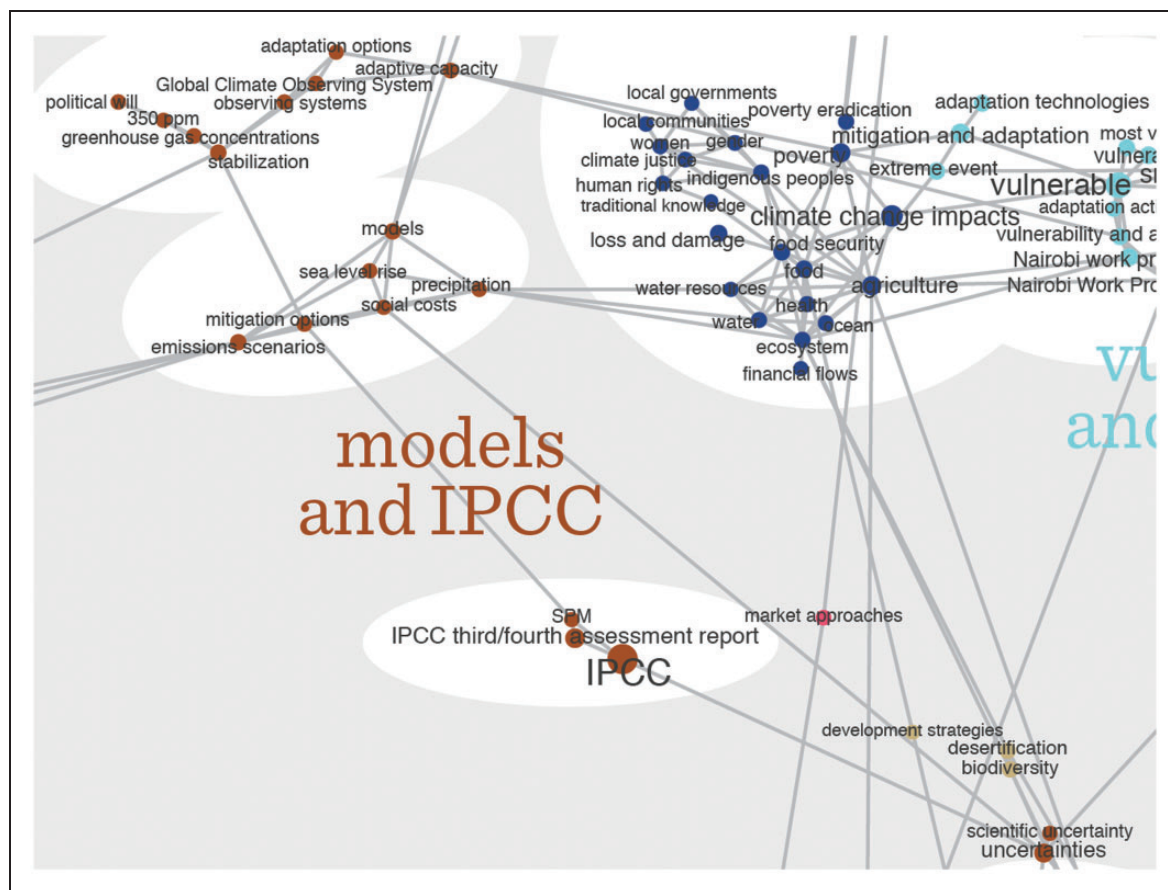


Figure 2. Zoom on the cluster “Models and IPCC.” The cluster contains terms related to the scientific assessment of climate change, and not surprisingly it is situated at the center of the network where it bridges the clusters on adaptation (to the right) and the clusters on mitigation (to left, top, and bottom). For more information on how this network has been designed, see the caption of Figure 1. A high-resolution and zoomable version of this image can be found at: <http://bds.sagepub.com/content/1/2/2053951714543804/F2.large.jpg>.

defining which *gases* should be reduced (CH_4 , CO_2 , N_2O , $HFCs$, etc.) and how (*bubble, gas-by-gas, all gases, inventories of GHGs emissions*). Second, there are two key terms whose specificity lies in their hybrid origin, a mix between science and politics: the *dangerous anthropogenic interferences with climate system* and the *levels of 1990* taken as a basis to set reducing targets. Third, there are political and negotiation terms such as *commitments for non-Annex I countries*, *new commitments*, and *common but differentiated responsibilities*.

The cluster configured around the term *Kyoto Protocol* (purple) is made of terms used for the implementation of the binding reduction mechanisms: first through the *Kyoto Protocol*, then through the definition of its successor. These negotiations, from the *Berlin Mandate* (1995) to the *Durban Platform* (2011), focus on various aspects: the definition of a *base year*, of *reference levels*, and of a *length of commitment period*; the implementation of *flexibility mechanisms*, among which are the *emission trading scheme*; the definition of *accounting rules* and *quantified emission limitation*

and *reduction objectives (QELROS)* and an agreement on the mechanisms for the transformation of pledges into quantified objectives (*pledges into QELROS*); the identification of possible means to address the *surplus Assigned Amount Units*. At the bottom of the cluster, terms related to the definition of *future commitments* after the end of the Kyoto Protocol (such as *second commitment period*, *gap between commitment periods* and *binding commitments*) connect this cluster to the *post-Kyoto* cluster.

Energies, technology transfer, and clean development projects: In the upper parts of the global negotiations map, the second mitigation arena is composed of three thematic clusters (Figure 4). On the left, we find two closely linked clusters, one with terms on reducing emissions from combustion of fossil fuels, especially in international transportation, and the other with terms on technology transfer. These two clusters are connected (toward the bottom) to the adaptation issues and in particular to the *impacts cluster* by the terms *development* and *sustainable development*.

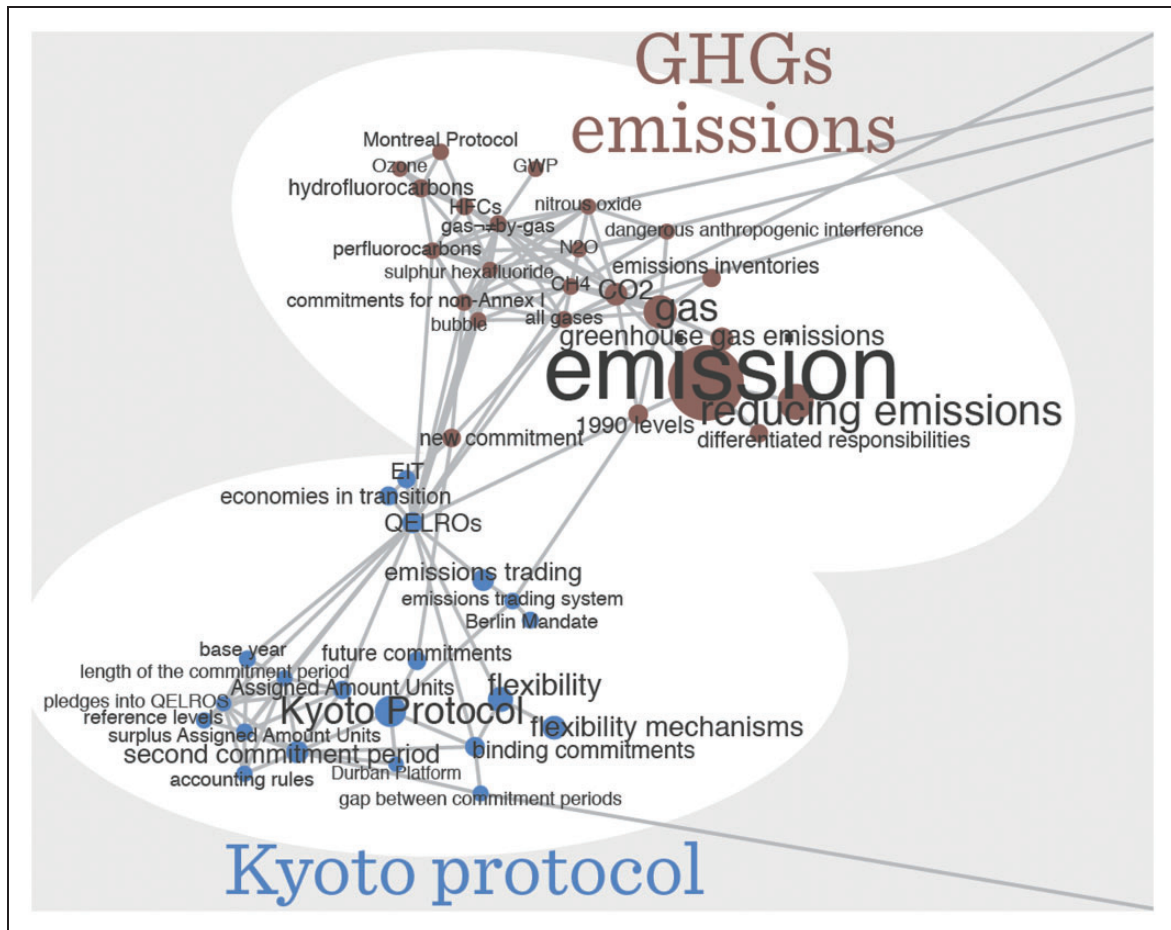


Figure 3. Zoom on the clusters “GHGs emissions” and “Kyoto Protocol.” The two clusters are tightly connected and contain terms related to the reduction of greenhouse gases and the Kyoto agreement. These clusters are connected to the scientific clusters presented in Figure 2 and to the discussion about the post-Kyoto mitigation agreement (see Figure 5). For more information on how this network has been designed see the caption of Figure 1. A high-resolution and zoomable version of this image can be found at: <http://bds.sagepub.com/content/11/2/2053951714543804/F3.large.jpg>.

On the far right corner of the map, the “CDM” refers to the flexibility mechanism of the Kyoto Protocol allowing a country to reduce its emissions implementing projects outside its territory (and in particular in developing countries). This mechanism fueled animated debates on two principles surrounding its implementation. On one hand, the principle of *supplementarity*: CDM projects are to remain supplemental to domestic efforts to reduce emissions. On the other hand, the principle of *additionality* requires that CDM projects really make additional reductions of GHGs emissions that would not have been achieved otherwise. The projects’ *impact assessments* are therefore crucial.

The management of carbon sinks: Since the Buenos Aires Plan of Action and the discussions on the definition of the implementation rules of the Kyoto Protocol, two new topics have arisen on the negotiation agenda (Figure 5).

First, discussions to reduce emissions caused by deforestation emerged in the late 1990s.

These discussions created real controversy, centered on debates about how to include LULUCF within Kyoto’s binding targets or offset mechanisms. Some countries insisted that this type of *carbon storage* should be excluded from the Kyoto Protocol because, in the absence of strict regulations, land management could become a pretext not to comply with emissions reduction. These regulations were defined with the Marrakech agreement setting a limit to the authorized LULUCF *carbon sink* deductions and leaving the possibility for *market-based* action to help Non-Annex I countries with *land use and forest management and conservation*. This agreement helped to ensure the ratification of the Kyoto Protocol in 2004.

Meanwhile, the Kyoto fairness principle (common but differentiated responsibilities) was put into question through the mobilization of the term *historical responsibility*, a formulation preferred by the BRIC countries. Not surprisingly, it is precisely this term that connects

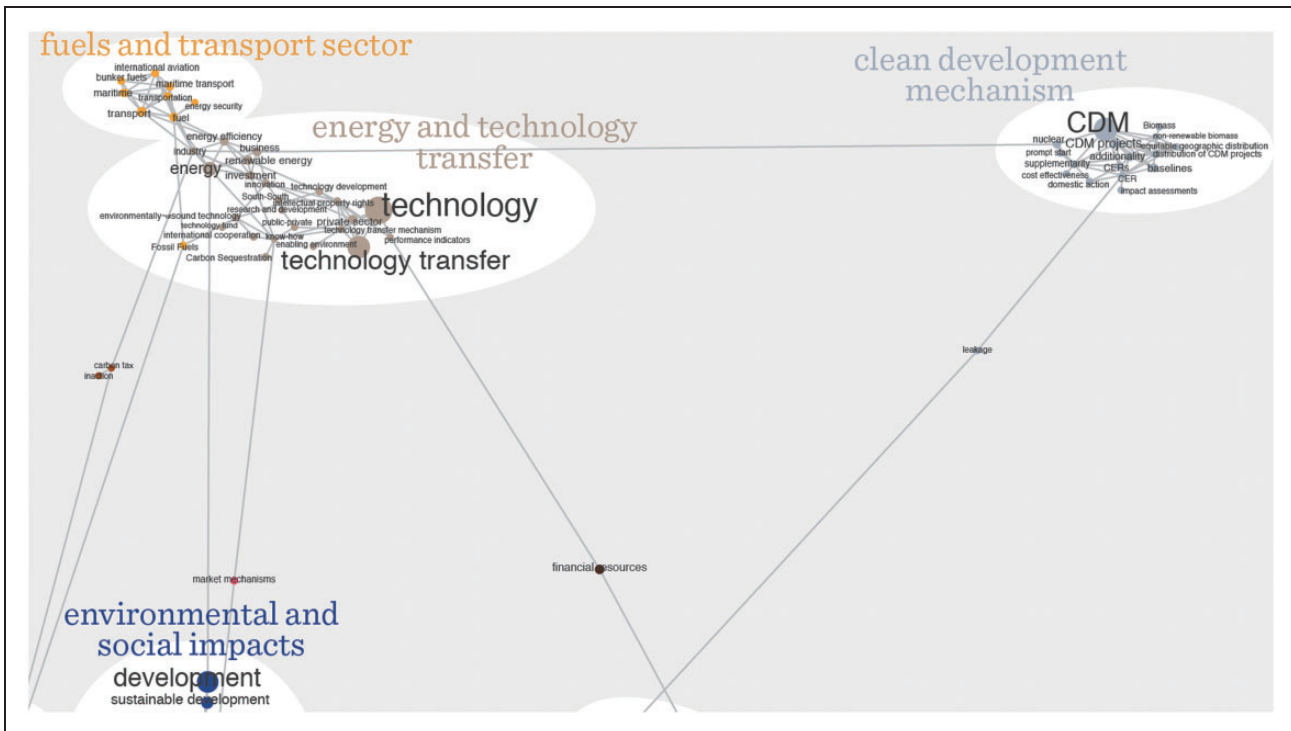


Figure 4. Zoom on the clusters “Fuels and transport sector,” “energy and technology transfer,” and “clean development mechanism.” Situated at the top of the network, the terms connected to energy and fuels are very tightly connected. Not far, but in a more marginal position, the cluster dedicated to the CDM contains terms related to the flexibility mechanisms of the Kyoto Protocol allowing states to reduce their emissions outside their national borders. For more information on how this network has been designed, see the caption of Figure 1. A high-resolution and zoomable version of this image can be found at: <http://bds.sagepub.com/content/1/2/2053951714543804/F4.large.jpg>.

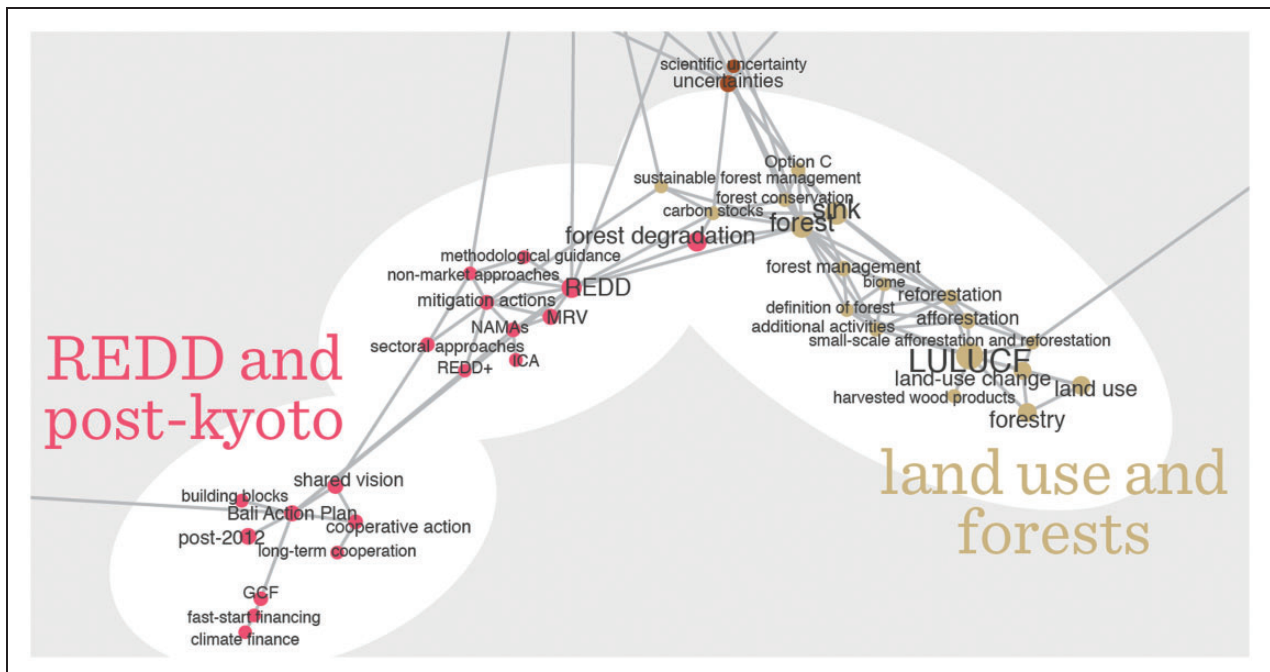


Figure 5. Zoom on the clusters “REDD and post-Kyoto” and “land use and forests.” Though referring to two thematically distinct negotiation arenas, the proximity of these two clusters reflects the importance that the question of agriculture and forestry has assumed in the latest years of the climate negotiations. For more information on how this network has been designed, see the caption of Figure 1. A high-resolution and zoomable version of this image can be found at: <http://bds.sagepub.com/content/1/2/2053951714543804/F5.large.jpg>.

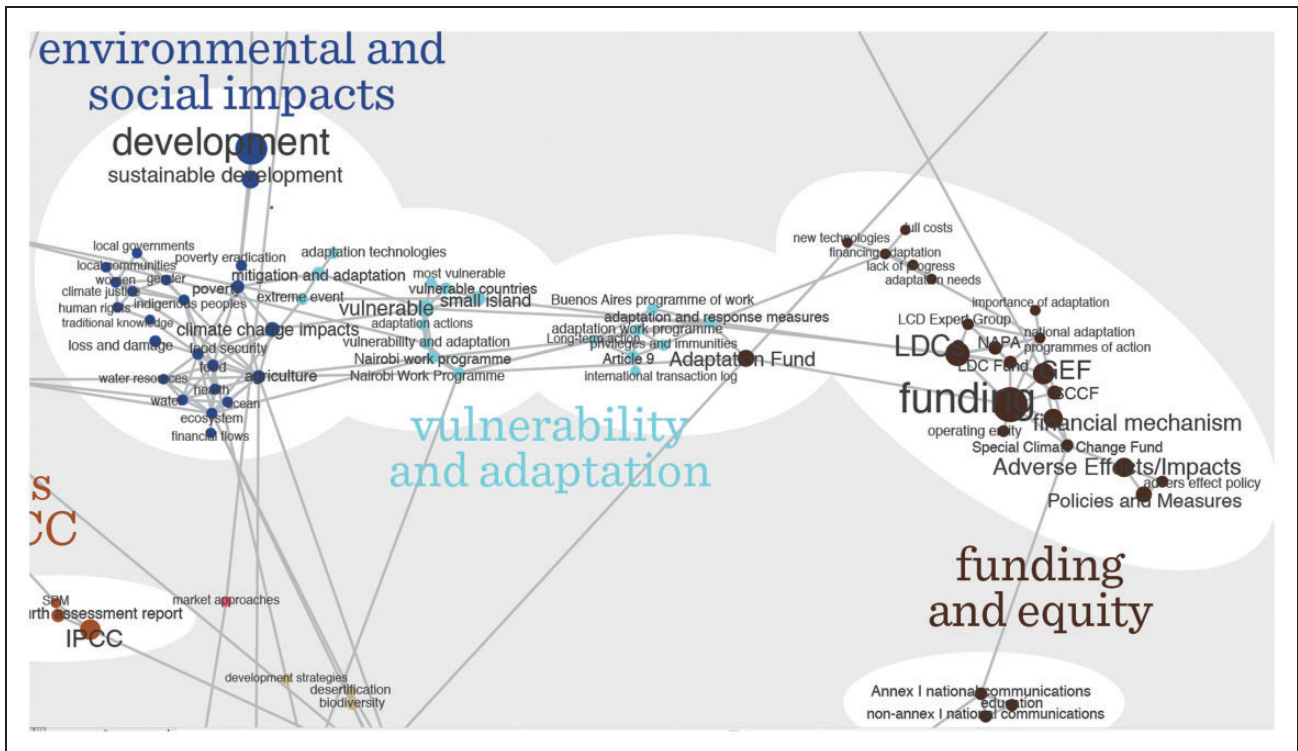


Figure 6. Zoom on the clusters “environmental and social impacts,” “vulnerability and adaptation,” and “funding and equity.” The three clusters contain the terms connected to the discussion about adaptation to climate change in the UNFCCC. Interestingly, these three clusters are tightly connected and located at the center of the network, suggesting that the debate about adaptation may be more “compact” and thematically coherent than the debate about mitigation (which is more present but also more dispersed in the network). For more information on how this network has been designed, see the caption of Figure 1. A high-resolution and zoomable version of this image can be found at: <http://bds.sagepub.com/content/11/2/2053951714543804/F6.large.jpg>.

the cluster on Kyoto Protocol to the cluster on post-Kyoto agreement, linking directly to the term *Bali Action Plan*. Adopted in 2007, such a plan defines the *building blocks*, of a post-Kyoto agreement: the strengthening of the mitigation policies, the setting up of adaptation policies to the impacts of climate change, the transfer of technology, as well as the definition of a *shared vision for long-term cooperative action*.

The Bali Action Plan also initiated discussions on how to reduce emissions caused by *deforestation and forest degradation (REDD)* as well as on the possibility for *sectoral approaches* to reducing emissions in developing countries.

The path to adaptation, impacts, and vulnerability. The center of the map is occupied by three clusters closely connected and dedicated to the issue of adaptation, environmental and social impacts, vulnerability and adaptation action, and adaptive funding and equity (Figure 6). Compared to the mitigation clusters, adaptation clusters are fewer (three against eight) and more compact. This shows the difference of the adaptation status in the UNFCCC negotiations. Where mitigation is the primary objective of the conference, and thus formulated in

numerous ways, adaptation, impacts, and vulnerability seem not only more limited in their articulation, but also more commonly connected to other themes (which accounts for their centrality in the map). For the sake of brevity (and in order to leave space for the other maps), we are saving the analysis of the adaptation, impacts, and vulnerability clusters for a forthcoming and more detailed article on the politics of adaptation in the negotiations.

Second map: Countries and COPs

Though missing on the first map, the temporal dimension is present in the next two diagrams.¹⁰ The second map shows the number of interventions by COP in the 21 most active countries in the negotiations (according to the ENB summaries).

In order to read the diagram properly, two things have to be remarked upon. First, the total number of paragraphs in each COP is not the same. At first glance, it is clear that the highest number of country occurrences appears during COP6 (The Hague) and COP15 (Copenhagen), which could mean that these negotiations were the most passionate. The Hague conference was a failure, leading to the organization of the

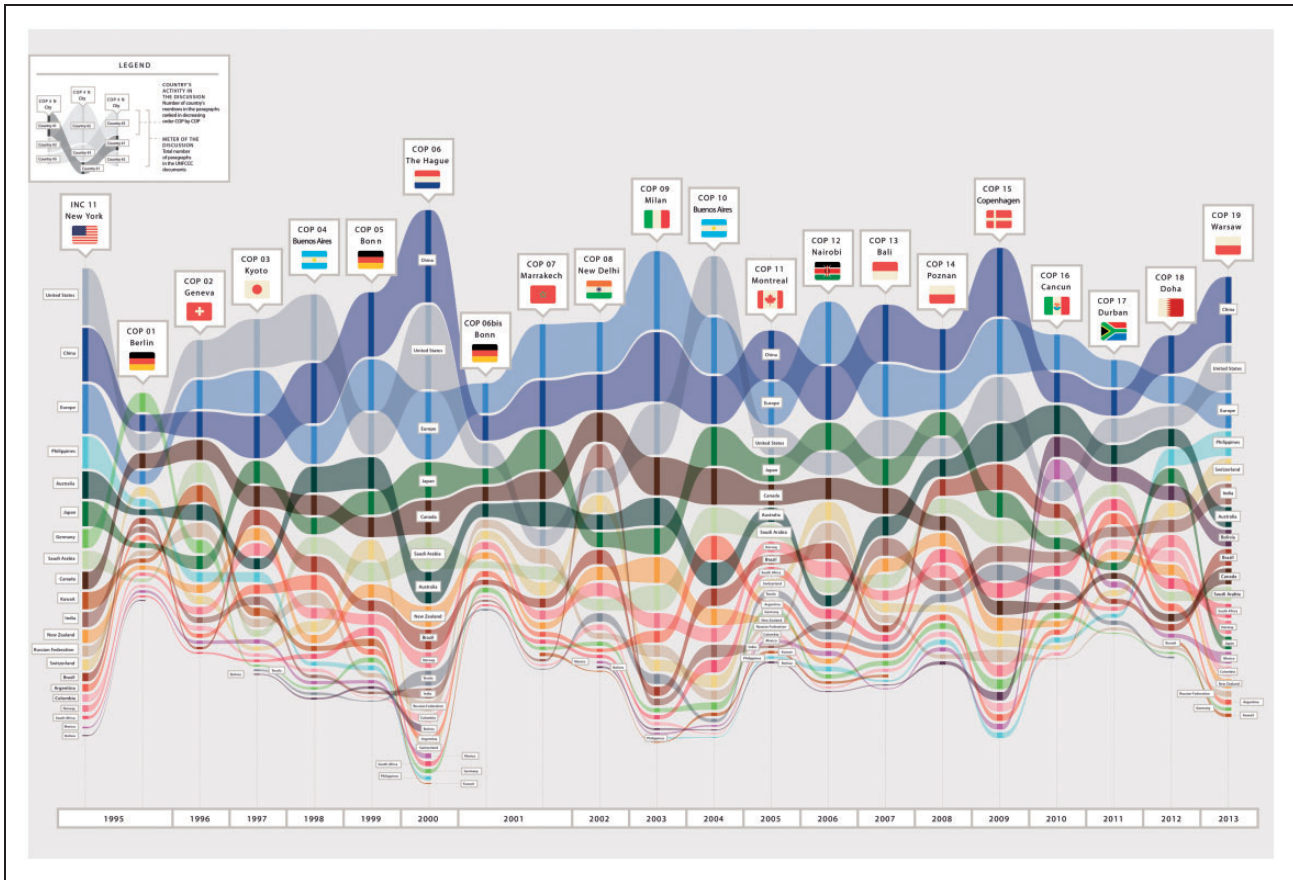


Figure 7. Streamgraph of the absolute and relative visibility of negotiation parties. The size of each country's flow is proportional to the number of paragraphs in which it is mentioned (the bigger the flow in any given COP, the more visible the country was at the time). Then, flows are sorted according to the number of occurrences: for each COP, the highest flow corresponds to the most active country while the lowest corresponds to the least active. For example, the United States is the most visible country in the first meeting and China in the last. A high-resolution and zoomable version of this image can be found at: <http://medialab.sciences-po.fr/publications/misunderstandings/> and <http://bds.sagepub.com/content/1/2/205395/1714543804/F7.large.jpg>.

COP6bis (Bonn) that same year, and COP15 performed well under expectations, failing to produce a post-Kyoto agreement. However, it is also possible that these two COPs were just covered in more detail by the staff of the ENB. For this reason, one can compare the size of different countries in one COP and the relative position of one flow through time, but one should not compare sizes over different COPs.

The second remark concerns the fact that while counting the number of occurrences of each country name, we did not distinguish between the sentences in which countries appear as subjects and the sentences in which they are mentioned as objects. However, we have noticed that in most cases the occurrence of a country name signals that the country has “taken the floor” in the negotiations. As a consequence, our measure indicates, strictly speaking, the visibility of a country in a COP, but it can also be read more largely as an indicator of activity.

In general, the diagram shows an outstanding stability: most countries stick to their relative position during

the 19 COPs. The 10 most active countries are represented by a rather stable, small group, which includes the United States, China, Europe, Australia, and Japan. The three leaders of the negotiations – China, the United States, and Europe – are ubiquitous and heading the negotiations. China, often speaking in the name of the G77 developing countries, never ranks beyond third position whereas Europe's position varies between the first and fifth ranks, and the United States between the first and seventh ranks. It can also be observed that countries tend to be more active when they host the negotiations: Germany is first in Berlin 1995, Japan is fourth in Kyoto 1997; India is fourth in New Delhi 2002; Canada is fifth in Montreal 2005.

Several exceptions should, however, be outlined. First, the Philippines and Bolivia, two countries from the southern hemisphere, have taken on very active roles, perhaps disproportionate with their size. Bolivia – very discreet during the first 15 COPs – has stood out

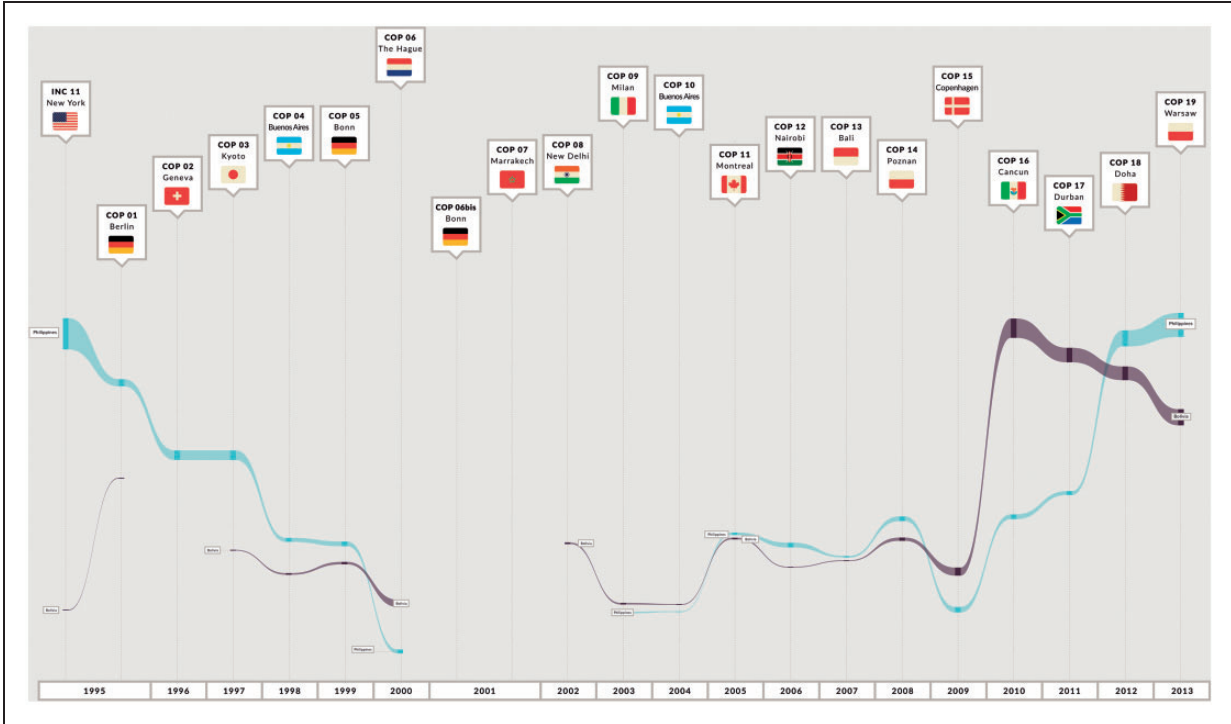


Figure 8. Figure 7 filtered to show only the flows of Bolivia (violet) and the Philippines (light blue), two countries remarkable for an increase in their visibility in the last UNFCCC's Conferences of Parties. For more information on how this diagram has been designed, see the caption of Figure 7. A high-resolution and zoomable version of this image can be found at: <http://medialab.sciences-po.fr/publications/misunderstandings/> and <http://bds.sagepub.com/content/1/2/2053951714543804/F8.large.jpg>.

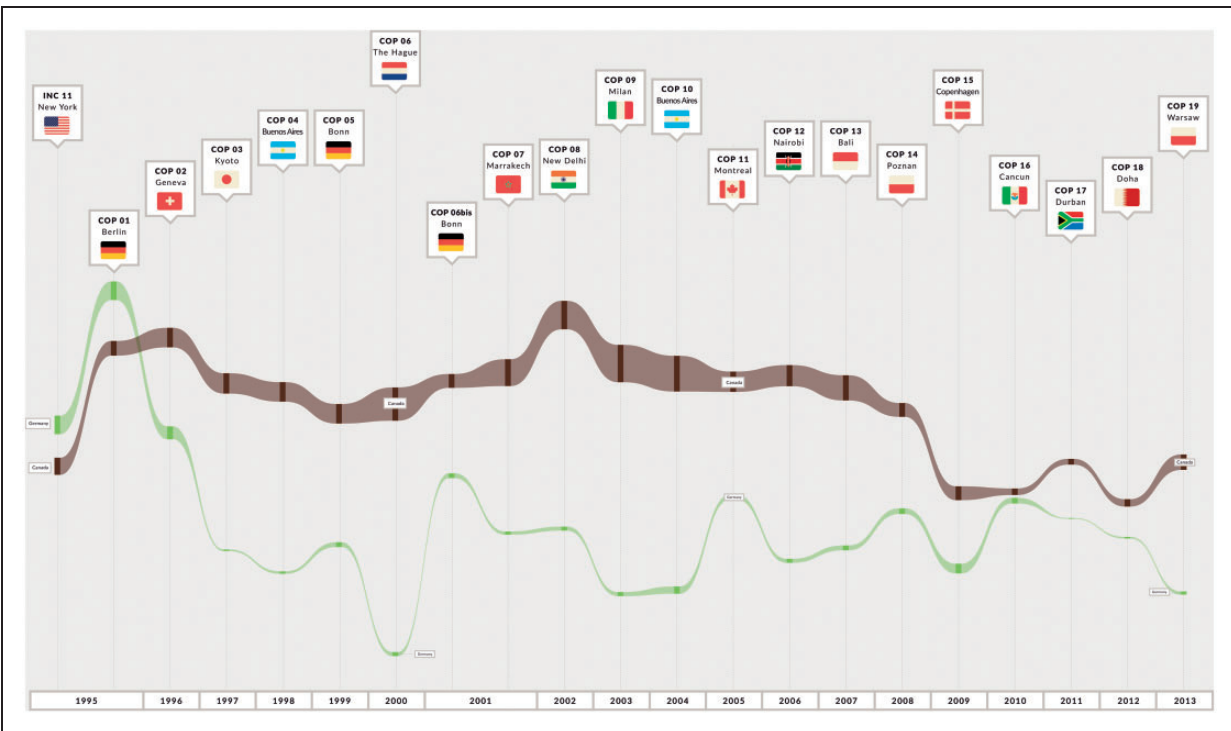


Figure 9. Figure 7 filtered to show only the flows of Canada (brown) and Germany (green). In contrast to the countries of Figure 8, these two countries have been chosen because their visibility decreases in the last years of the climate negotiations. If the case of Germany can be explained by the decision of the European states to negotiate as a group rather than individually, the decrease in Canada's visibility may suggest a progressive disengagement of the country from the issue of climate change. For more information on how this diagram has been designed, see the caption of Figure 7. A high-resolution and zoomable version of this image can be found at: <http://medialab.sciences-po.fr/publications/misunderstandings/> and <http://bds.sagepub.com/content/1/2/2053951714543804/F9.large.jpg>.

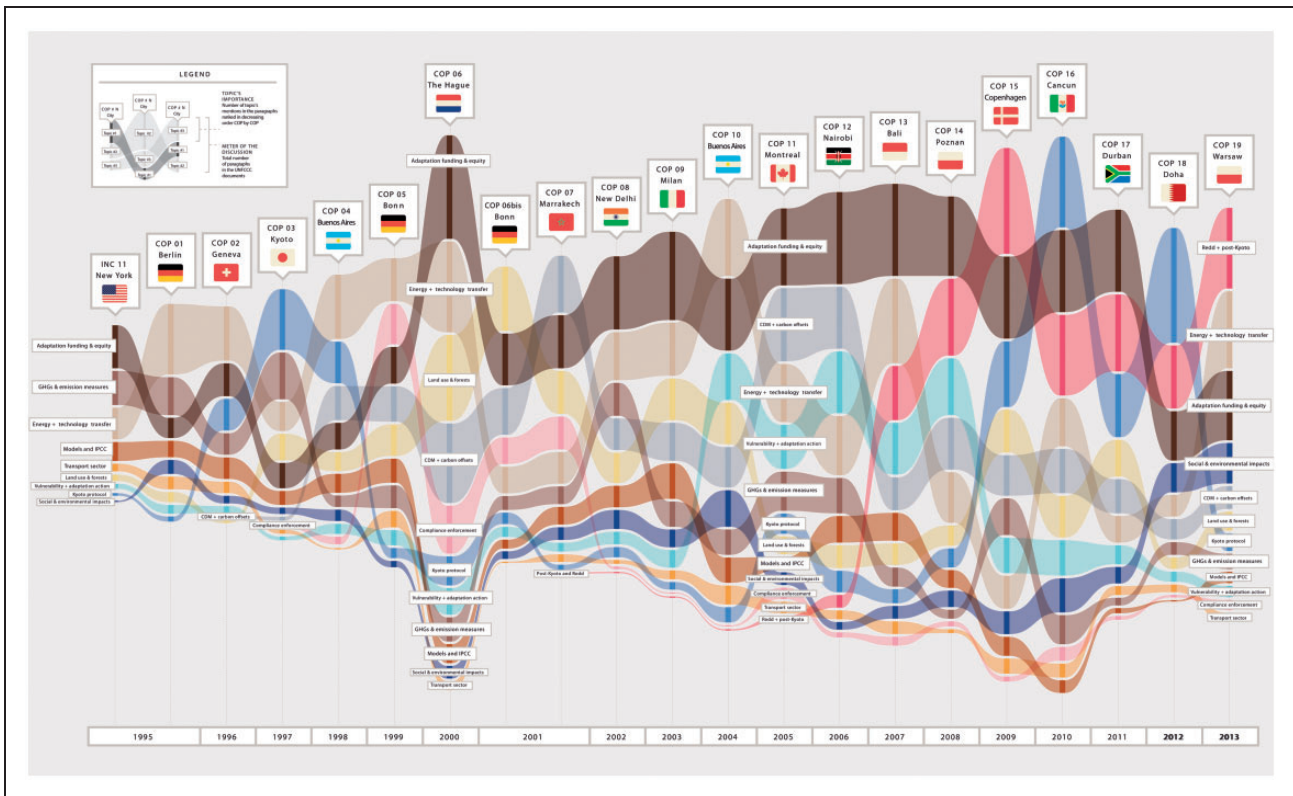


Figure 10. Streamgraph of the absolute and relative visibility of the themes of the negotiation. The size of each theme flow is proportional to the number of paragraphs in which the two terms defining the themes are present. Then, flows are sorted according to the number of occurrences: for each COP, the highest flow corresponds to the most visible theme while the lowest corresponds to the least visible. A high-resolution and zoomable version of this image can be found at: <http://medialab.sciences-po.fr/publications/misunderstandings/> and <http://bds.sagepub.com/content/1/2/2053951714543804/F10.large.jpg>.

from COP16 (Cancun) onwards, and has been one of the leading voices around “loss and damages.” Bolivia often comments on issues related to the historical responsibility of developed countries and their compliance with their commitments to reduce GHGs emissions.

The Philippines’ trajectory is also interesting: quite conspicuous in the early negotiations (fourth rank at the INC11 in New York and sixth rank at the COP1 in Berlin), the country steps aside during the next conferences to stand out again in Doha (COP18) and Warsaw (COP19). If the Philippines mainly speaks out on equity and “common but differentiated responsibilities” – principle 7 of the Rio Declaration on Environment and Development – and on funding and adaptation funds, Doha and Warsaw conferences have witnessed many references to the two “unprecedented” typhoons that devastated the Philippines (Bopha/Haiyan) at that very time.

The visibility of some countries increases in a punctuated fashion at very specific COPs. Mexico, for example, shows a rather low profile during most negotiations, but ranks 5 during COP16 (Cancun), organized in Mexico. Tuvalu’s trajectory bears

mentioning as well: from the Kyoto conference onwards, this small Pacific island has ranked among the 21 most visible member countries. Yet, Tuvalu also reached rank 13 in Poznan (COP14), rank 19 in Copenhagen (COP15), and rank 12 in Cancun (COP16). During these conferences, Tuvalu mainly addressed the issue of a successor to the Kyoto Protocol – the island even supports its own protocol proposition.

We can also testify to the withdrawal of Canada from climate negotiations. Canada ranks among the six more visible countries until COP13 in Bali. It then withdraws from the ranks of top participating countries at the Poznan conference in 2008. By way of explanation, in 2006 a new conservative Prime Minister, Stephen Harper, was elected to lead Canada and in 2011 the country withdrew from the Kyoto Protocol and actively initiated unconventional oil drillings in the Athabasca region of Alberta. Germany is also less visible after COP1 organized in Berlin. The reason might be the increasing importance of the European Union as a representative of its Member States during the negotiations.

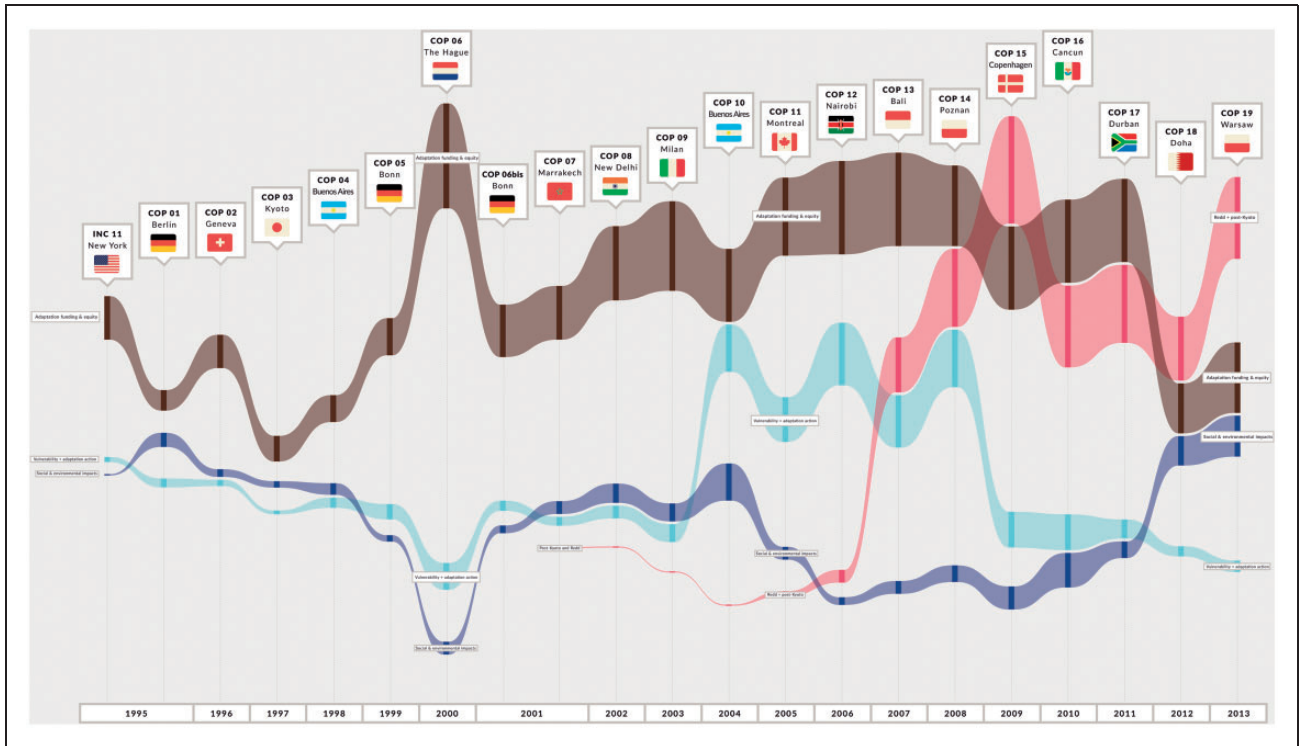


Figure 11. Figure 10 filtered to show only the flows of funding and equity (brown), vulnerability and adaptation (light blue), social and environmental impacts (dark blue), REDD and post-Kyoto (pink). This figure shows the evolution of the visibility of the three main clusters of terms related to adaptation. Whereas the debate on funding (in brown) is present throughout all the negotiations, the debate on the impacts of global warming (in dark blue) seems to gain visibility only in the last years of negotiation. The question of vulnerability (in light blue) seems to be highly discussed in the central years of the negotiations, but to lose visibility when the question of the post-Kyoto agreements (in pink) becomes central. For more information on how this diagram has been designed, see the caption of Figure 10. A high-resolution and zoomable version of this image can be found at: <http://medialab.sciences-po.fr/publications/misunderstandings/> and <http://bds.sagepub.com/content/1/2/2053951714543804/F11.large.jpg>.

Third map: Issues and COPs

The third map is built with the same visualization process as the second one, but the flows represent the topics (not the countries) in climate negotiations. The topics correspond to the clusters identified in the first map and the computation of their visibility is made by counting the number of paragraphs in which at least two words of the same cluster are present.

Whereas adaptation and mitigation issues are both central at the UNFCCC, mitigation has always been a top priority on the negotiations' agenda from the very beginning. During the UNFCCC's first negotiations, the focus was on the responsibility of developed countries and their effort to reduce their GHGs emissions. Discussions were initiated in Berlin (1995), setting up a restrictive framework to reduce emissions of the Annex I countries, an objective that was achieved in 1997 with the adoption of the Kyoto Protocol.

In this first phase of the negotiations, little attention was dedicated to the actions of developing countries to cope with the impacts of climate change, with a notable

exception. From the very beginning, the most vulnerable members have succeeded in putting on the agenda the issue of the financing of adaptation actions. The question of adaptation provisions was included in the Kyoto Protocol (1995) and generated a tense debate on the mechanisms to ensure this financing at The Hague (COP6) with members disagreeing on the types of financing, modalities, and competencies.

Adaptation, however, assumed a greater importance in the second phase of the negotiations. With all parties facing difficulties in achieving their mitigation objectives, debates on what shall be done regarding vulnerability, climate change impacts, and adaptation, as well as how to finance these actions, increased between Marrakech (2000) and Bali (2007) conferences, to become a matter of general concern. The COPs of New Delhi (2002) and Buenos Aires (2004) are often referred to as "the adaptation COPs." Following a series of extreme weather events that affected both developing and developed countries, the COP 11 in Montreal (2005) marks the end of the illusion of invulnerability of developed countries. Adaptation assumed

therefore a central place in the negotiation as recognized by the four pillars of the Bali Action Plan (2007), which include adaptation as well as mitigation, technology transfer, and financing within a perspective of long-term coordinated action. Since the Poznan conference (COP14), adaptation seems definitively established as a subject of climate diplomacy and is therefore less discussed.

The discussions in Nairobi (2006) evolved toward the operationalization of adaptation funds while the consequences of climate change became more apparent. In the last COPs, the debate around adaptation shares increasing “floor time” with a renewed interest in carbon sinks (forest and agriculture), raised by the United States, and the debate on how to help developing countries achieve their objectives in terms of emission reduction. In this context, land-use mechanisms, land-use change, and forestry (LULUCF), CDM and compliance projects are strongly debated. An agreement was reached in Buenos Aires (2004) and Bali (2007) and the debate on these issues, as well as on technology transfers in order to reduce emissions of fossil fuels, is reaching a gradual stabilization.

Then, and especially after the Copenhagen failure, mitigation of industrial emission sources comes back at the top of the negotiating agenda because of the difficulties of finding a successor to the Kyoto Protocol and the challenges in implementing REDD and other developing country emission reduction initiatives. Adaptation discussion remains, however, important because of the increasing acknowledgement of the social impacts of climate change and the proposal of developing countries to establish a financial mechanism to cover the “loss and damages” that they may suffer because of global warming.

Conclusion

In this paper, we have discussed a digital analysis of a corpus of reports on the international negotiations on climate change. Though such an investigation is still in progress, the analysis has allowed us to highlight three common difficulties in digital research and to obtain three preliminary visualizations. The two results are of course connected. As we have tried to show, digital research can only be productive when scholars:

1. Draw on the appropriate corpus of digital traces (the first misunderstanding would have led us to search traces on climate diplomacy in generic datasets such as the scientific literature or the social media).
2. Take into consideration the conditions of production of such traces (the second misunderstanding would have led us to mistake our proxy on the negotiations – the ENB – for the phenomenon itself,

thereby disregarding the specific mediations operated by such a proxy and its specific format).

3. Are not afraid to make the choices and the efforts necessary to clean, transform, visualize, and interpret the data that they have collected (the third misunderstanding would have dissuaded us from using what we know to mine interesting findings out of the overabundance of variables and correlations).

By gathering a corpus of traces specifically focused on climate negotiations, by investigating and exploiting the characteristic nature of such traces, and by using our expertise of climate diplomacy to cultivate the emergence of interesting findings, we have obtained three interesting results. First, we have been able to identify the thematic clustering of the discussions in the UNFCCC and the articulation among the different issues. Second, we have been able to track the visibility of different countries across 18 years and 21 international meetings. Third, we have visualized the rise and fall of the different themes of climate diplomacy.

These results are promising because they are both expected and surprising. In developing new digital methods for the social sciences, we find ourselves confronted by the classic problem of the “experimenter regress.” As Harry Collins (1975) observed in the case of relativistic physics, the problem of original research is that, since both its methods and its theories are tentative, it is hard to find a stable ground to establish its validity. This is also true in the case of digital research. Since this research intends to renew social theory through a series of novel research methods, its claims are still difficult to ground: both its conceptual and its methodological tools remain uncertain. The only way to bootstrap digital social sciences out of the experimenter regress is to compare its results with those obtained in traditional research and hope that our findings are consistent enough with previous knowledge to be credible and yet original enough to provide new insights.

To some extent, this is what we obtained with the three maps we just presented. The first and the third maps illustrated, as expected, the preeminent role in climate diplomacy of the questions related to mitigation. Mitigation constitutes the bulk of UNFCCC’s discussions. Its different sub-issues (the measure of GHGs, the technology transfer, the transports, the CDM, the carbon sinks in land and forests) spread all over and throughout all the negotiations. Mitigation articulates the space of the debate (infusing the discussions on GHGs reduction, encouraging energy transition, and defending the carbon sinks) and defines its rhythm (with the fluctuation of the debates about a binding protocol).

Adaptation, on the other hand, appears as a specific topic of the negotiations: a tightly connected group of issues located in a precise position on the map. Yet, and

this was not obvious before our analysis, adaptation appears to occupy the center of the climate negotiations and to have been present and highly visible from the very beginning of the negotiations. These findings stand out against the many claims in the literature about climate diplomacy as to an “adaptation turn” in the last years of the negotiation (Howard, 2009; Pielke et al., 2007). But comparing closely maps 1 and 3, an interesting explanation emerges. What has always been present and visible in the negotiations is not the whole discussion about adaptation but the specific question of adaptation finance. Interestingly, this question appears to be the most marginal of the adaptation-related topics, with a position that is not structurally different from that of the topics of mitigation. An “adaptation turn,” however, can be recognized in the rise of the question of vulnerability (from COP9 to COP14) and in the more recent ascent of the question of the climate impacts (from COP15). These are the two clusters that occupy the center of map 1. Reading the two maps together, therefore, the hypothesis can be made that, in the last 10 years, the emergence of the first recognizable effects of climate change has gradually occupied the center of the negotiation scene, not as much replacing the previous discussion but somehow bridging discussions that would have been separated otherwise. This hypothesis, to be sure, needs to be confirmed by further analysis (which we are currently carrying out), but this little example has at least showed that, when performed properly, digital research can produce original and consistent results.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Funding

This publication results from research undertaken in the framework of the MEDEA project (Mapping environmental debates on adaptation) funded by the French National Research Agency (ANR) (reference ANR-11-CEPL-0004).

Notes

1. United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro.
2. This is Article 1 of the Convention, which can be found online at: unfccc.int/files/essential_background/background_publications_htmlpdf/application/pdf/conveng.pdf
3. The entire text of the *Earth Negotiations Bulletin*, Volume 12, is available at: <http://www.iisd.ca/vol12/>.
4. On the invention of “representativity” in classical statistics, cf. Desrosières, 1991 (in particular pp. 211–213).
5. See Hedman et al. for a more detailed repertoire of the work on data in digital social sciences. In another field (climatology), cf. Paul Edwards’ (2010) discussion of the construction, turnover and continual transformation of data in climate modeling (particularly chapters 8 and 10).
6. Cf. <http://www.iisd.ca/about/staff.htm>, <http://www.iisd.ca/about/équipe/>
7. Cf. en.wikipedia.org/wiki/AOL_search_data_scandal
8. Since 2005, COPs have been held simultaneously with MOPs (Meeting of Parties of the Kyoto Protocol), whose report is included in the same document.
9. The work described in this article has become the basis for a more systematic initiative to analyze the debate in the UNFCCC. This initiative is meant to create an online platform to map the discussion of COP21 in 2015 in Paris. Such platform should provide the actors of the debate (negotiators, NGOs, lobbies, journalists) as well as the general public with a better grasp on the complexity of climate negotiations.
10. All following figures (from Figures 7 to 11) have been designed by Martina Elisa Cecchi, whom we thank a lot for her extraordinary work.

References

- Anderegg WRL, Prall JW, Harold J, et al. (2010) Expert credibility in climate change. *Proceedings National Academy Sciences United States of America* 107(27): 12107–12109. DOI:10.1073/pnas.1003187107.
- Barnes TJ (2014) What’s old is new, and new is old: History and geography’s quantitative revolutions. *Dialogues in Human Geography* 4(1): 50–53. DOI:10.1177/2043820614525707.
- Biber D (1993) Representativeness in Corpus Design. *Literary and Linguistic Computing* 8(4): 243–257. DOI:10.1093/lc/8.4.243.
- Björström A and Polk M (2011) Physical and economic bias in climate change research: A scientometric study of IPCC Third Assessment Report. *Climatic Change* 108(1–2): 1–22. DOI:10.1007/s10584-011-0018-8.
- Blondel VD, Guillaume J-L, Lambiotte R, et al. (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics* 2008(10): 10008.
- Bollier D and Firestone C (2010) *The Promise and Peril of Big Data*. Washington: The Aspen Institute. Available at: www.ilmresource.com/collateral/analyst-reports/10334-ar-promise-peril-of-big-data.pdf (accessed 5 July 2014).
- Boyd D and Crawford K (2011) Six provocations for Big Data. *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011. Available at: <http://ssrn.com/abstract=1926431> or <http://dx.doi.org/10.2139/ssrn.1926431> (accessed 5 July 2014).
- Carr N (2014) The limits of social engineering. *MIT Technology Review*. Available at: www.technologyreview.com/review/526561/the-limits-of-social-engineering/ (accessed 5 July 2014).
- Cresswell T (2014) Deja vu all over again: Spatial science, quantitative revolutions and the culture of numbers. *Dialogues in Human Geography* 4(1): 54–58. DOI:10.1177/2043820614525715.
- Collins HM (1975) The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in

- physics. *Sociology* 9: 205–224. DOI:10.1177/003803857500900202.
- Couper M (2000) Web surveys: A review of issues and approaches. *Public Opinion Quarterly* 64(4): 464–494. Available at: www.ncbi.nlm.nih.gov/pubmed/11171027 (accessed 5 July 2014).
- Dalton C and Thatcher J (2014) What does a critical data studies look like, and why do we care? Seven points for a critical approach to “big data”. *Society and Space Open Site*. Available at: <http://societyandspace.com/material/commentaries/craig-dalton-and-jim-thatcher-what-does-a-critical-data-studies-look-like-and-why-do-we-care-seven-points-for-a-critical-approach-to-big-data/> (accessed 5 July 2014).
- Desrosières A (1991) How to make things which hold together: Social science, statistics and the state. In: Wagner P, Witrock B and Whitley R (eds) *Discourses on Society*. Netherlands: Springer, pp. 195–218. DOI:10.1007/978-0-585-29174-1_8.
- Edwards P (2010) *A Vast Machine: Computer Models, Climate Data and the Politics of Global Warming*. Cambridge, MA: MIT Press.
- Eisenstein E (1979) *The Printing Press as an Agent of Change*. Cambridge: Cambridge University Press.
- Ess C and AoIR ethics working committee. (2002) Ethical decision-making and Internet research. Recommendations from the AOIR ethics working committee, pp. 1–33. Available at: www.aoir.org/reports/ethics.pdf.
- Gayo-Avello D (2012) “I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper”—A Balanced Survey on Election Prediction using Twitter Data. *arXiv Preprint arXiv:1204.6441*: 1–13. Available at: <http://arxiv.org/abs/1204.6441> (accessed 5 July 2014).
- Giles J (2012) Making the links. *Nature* 488: 448–450.
- Harford T (2014) Big data: Are we making a big mistake? *Financial Times Magazine*. Available at: www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2ysdIXgD2 (accessed 5 July 2014).
- Howard J (2009) Climate change mitigation and adaptation in developed nations: A critical perspective on the adaptation turn in urban climate planning. In: Davoudi S, Crawford J and Mehmood A (eds) *Planning for Climate Change: Strategies for Mitigation and Adaptation for Spatial Planners*. Sterling, VA: Earthscan, pp. 19–32.
- Jacomy M, Venturini T, Heymann S, et al. (forthcoming) ForceAtlas2, a graph layout algorithm for handy network visualization. *Plos ONE*. Available at: www.medialab.sciences-Po.fr/fr/publications-Fr/ (accessed 5 July 2014).
- Kleinberg J (2008) The convergence of social and technological networks. *Communications of the ACM* 51(11): 66. DOI:10.1145/1400214.1400232.
- Latour B (1995) The “Pédofil” of Boa Vista: A Photo-Philosophical Montage. *Common Knowledge*.
- Latour B, Jensen P, Venturini T, et al. (2012) “The Whole is Always Smaller Than Its Parts” A Digital Test of Gabriel Tarde’s Monads. *British Journal of Sociology* 63(4): 591–615.
- Law J, Ruppert E and Savage M (2011) The double social life of methods. Available at: [www.cresc.ac.uk/sites/default/files/The Double Social Life of Methods CRESC Working Paper 95.pdf](http://www.cresc.ac.uk/sites/default/files/The%20Double%20Social%20Life%20of%20Methods%20CRESC%20Working%20Paper%2095.pdf) (accessed 5 July 2014).
- Lazer D, Kennedy R, King G, et al. (2014) Big data. The parable of Google Flu: Traps in big data analysis. *Science* 343(6176): 1203–1205. DOI:10.1126/science.1248506.
- Lazer D, Pentland A, Adamic L, et al. (2009) Computational social science. *Science* 323(5915): 721–723. DOI:10.1126/science.1167742.
- Li J, Wang M-H and Ho Y-S (2011) Trends in research on global climate change: A science citation index expanded-based analysis. *Global and Planetary Change* 77(1–2): 13–20. DOI:10.1016/j.gloplacha.2011.02.005.
- Manovich L (2012) Trending: The promises and the challenges of big social data. In: Gold MK (ed.) *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, pp. 460–475.
- Marcus G and Davis E (2014) Eight (No, Nine!) Problems With Big Data. *The New York Times*. Available at: www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?_r=4 (accessed 5 July 2014).
- Marres N (2012) The redistribution of methods: On intervention in digital social research, broadly conceived. *The Sociological Review* 60: 139–165. DOI:10.1111/j.1467-954X.2012.02121.x.
- Niederer S (2013) “Global warming is not a crisis!”: Studying climate change skepticism on the Web. *NECSUS European Journal of Media Studies* 2: 83–112. DOI:10.5117/NECSUS2013.1.NIED.
- Pielke R Jr, Prins G and Rayner S (2007) “Climate change 2007: Lifting the taboo on adaptation”. *Nature* 445: 597–598.
- Reigeluth T (2014) Article Why data is not enough: Digital traces as control of self and self-control. *Surveillance and Society* 12(2): 243–254.
- Rieder B and Röhle T (2012) Digital methods: Five challenges. In: Berry DM (ed.) *Understanding Digital Humanities*. Houndmills: Palgrave Macmillan, pp. 67–84.
- Rogers R (2013) *Digital Methods*. Cambridge, MA: MIT Press.
- Rogers R and Marres N (2000) Landscaping climate change: A mapping technique for understanding science and technology debates on the World Wide Web. *Public Understanding of Science* 9: 141–163. DOI:10.1088/0963-6625/9/2/304.
- Savage M and Burrows R (2007) The coming crisis of empirical sociology. *Sociology* 41(5): 885–899. DOI:10.1177/0038038507080443.
- Stanhill G (2001) The growth of climate change science: A scientometric study. *Climatic Change* 48(2–3): 515–524. DOI:10.1023/A:1010721600896.
- Tabor A (2012) Making World Bank Open Data Usable: An Important Next Step. *My Science Work*. Available at: www.mysciencework.com/news/6556/making-world-bank-open-data-usable-an-important-next-step (accessed 5 July 2014).
- Uprichard E (2012) Being stuck in (live) time: The sticky sociological imagination. *The Sociological Review* 60: 124–138. DOI:10.1111/j.1467-954X.2012.002120.x.

- Uprichard E (2013) Big Data, Little Questions? *Discover Society*. Available at: www.discoversociety.org/focus-big-data-little-questions/ (accessed 5 July 2014).
- Venturini T and Latour B (2010) The social fabric: Digital traces and quali-quantitative methods. In: *Proceedings of Futur en Seine 2009*. Paris, pp.87–101. Paris: Editions Futur en Seine. Available at: www.tommasoventurini.it/web/uploads/tommaso_venturini/TheSocialFabric.pdf (accessed 5 July 2014).
- Viegas FB, Wattenberg M, Kriss J, et al. (2007) Talk before you type: Coordination in Wikipedia. In: *System Sciences, HICSS 2007. 40th Annual Hawaii International Conference*, p. 78. Hawaii: IEEE. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4076527.
- Watts DJ (2007) A twenty-first century science. *Nature* 445(7127): 489. DOI:10.1038/445489a.
- Weeds J and Weir D (2005) Co-occurrence retrieval: A flexible framework for lexical distributional similarity. *Computational Linguistics* 31(4): 439–475.